AD-A056 792 NATIONAL RANGE OPERATIONS DIRECTORATE WHITE SANDS MIS--ETC F/G 12/1
ROBUST REGRESSION: COMPUTATIONAL METHODS FOR M-ESTIMATES.(U)
JUN 78 W S AGEE, R H TURNER

UNCLASSIFIED ACD-TR-66 NL

1 OF 1

AD
A056792

END
DATE
FILMED
9-78
DDC

TECHNICAL REPORT

NO. 66

JUNE 1978

ROBUST REGRESSION:
COMPUTATIONAL METHODS FOR M-ESTIMATES

MATHEMATICAL SERVICES BRANCH
ANALYSIS & COMPUTATION DIVISION
US ARMY WHITE SANDS MISSILE RANGE
WHITE SANDS MISSILE RANGE, NEW MEXICO

TECHNICAL REPORT

NO. 66

Prepared by William S Agee
WILLIAM S. AGEE
Mathematician, Math Svcs Br

Robert H. Turner
ROBERT H. TURNER
Mathematician, Math Svcs Br

Reviewed by Jon E Gibson
JON E. GIBSON
Chief, Math Svcs Br

Approved by Patrick J. Higgins
PATRICK J. HIGGINS
Chief, Anal & Cmpt Div

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Analysis & Computation Division ✓ National Range Operations Directorate White Sands Missile Range, New Mexico 88002	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP NA

~~3. REPORT TITLE~~

ROBUST REGRESSION: COMPUTATIONAL METHODS FOR M-ESTIMATES (6)

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* (9) Technical rept.

5. AUTHOR(S) *(First name, middle initial, last name)*

William S./Agee [redacted] Robert H./Turner (10)

(12) 54 P.

~~6. REPORT DATE~~ JUN [redacted] 78 (11)	7a. TOTAL NO. OF PAGES 50	7b. NO. OF REFS 14

~~8a.~~ CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	~~9a. ORIGINATOR'S REPORT NUMBER(S)~~ (14) ACD-TR-66 ~~Technical Report No. 66~~ ✓ 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Psi

The computation of robust M-estimates of regression is considered in detail using the ψ functions of Huber, Andrews, and Hampel. The computation of M-estimates of regression is considered for linear models, linear models with vector observations, and nonlinear models. Examples are given using actual data for each of these different classes of models. Careful attention is given to the important problem of convergence of M-estimates with redescending ψ functions. A lengthy treatment of this problem is given for the Daniel and Wood data by considering several starting methods for the iterative solution and different breakpoints for the ψ functions.

Psi

447766

UNCLASSIFIED
Security Classification

kc

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Robust Regression M-estimates Outliers Data Reduction						

ABSTRACT

The computation of robust M-estimates of regression is considered in detail using the ψ functions of Huber, Andrews, and Hampel. The computation of M-estimates of regression is considered for linear models, linear models with vector observations, and nonlinear models. Examples are given using actual data for each of these different classes of models. Careful attention is given to the important problem of convergence of M-estimates with redescending ψ functions. A lengthy treatment of this problem is given for the Daniel and Wood data by considering several starting methods for the iterative solution and different breakpoints for the ψ functions.

TABLE OF CONTENTS

1. INTRODUCTION

The estimation of coefficients in a linear regression model by least squares has long been plagued by the possible presence of outliers, i.e., observations which for some reason do not belong with the major portion of the observations or with the regression model. To quote Huber [1], "even a single grossly outlying observation may spoil the least squares estimate and moreover outliers are much harder to spot in the regression case than in the simple location case." Several robust alternatives to the use of least squares in estimating the coefficients in a linear regression model have been developed which are outlier resistant. Robust statistical methods may be loosely described as those which will perform well under a variety of underlying distributions or in the presence of observations from contaminating distributions.

Robust estimation methods have been classified by Huber [1] and [2]. Huber's classifications are termed L-estimates, M-estimates, and R-estimates. The L-estimates are formed as linear combinations of the order statistics. The α - trimmed mean is an example of an L-estimate for a location parameter. The R-estimates are derived on the basis of rank tests. The estimate of location obtained by taking the median of all pairwise averages of the observations is an R-estimate. Probably, the most popular robust regression methods are the M-estimates. Their popularity stems from their generality, their close computational relationship to least squares, and the ease of numerical computation.

2. M-Estimates for Regression

Given the linear model

$$y_i = \sum_{j=1}^{p} x_{ij}\theta_j + e_i \qquad i=1,n \tag{1}$$

where the regression parameters θ_j are unknown and to be estimated from a knowledge of the values y_i and x_{ij}. The M-estimates of θ_j minimize

$$\sum_{i=1}^{n} \rho\left(y_i - \sum_{j=1}^{p} x_{ij}\theta_j\right) \tag{2}$$

where $\rho(\cdot)$ is some suitable function. Differentiating (2) leads to

$$\sum_{i=1}^{n} X_i^T \Psi(y_i - x_i\theta) = 0 \tag{3}$$

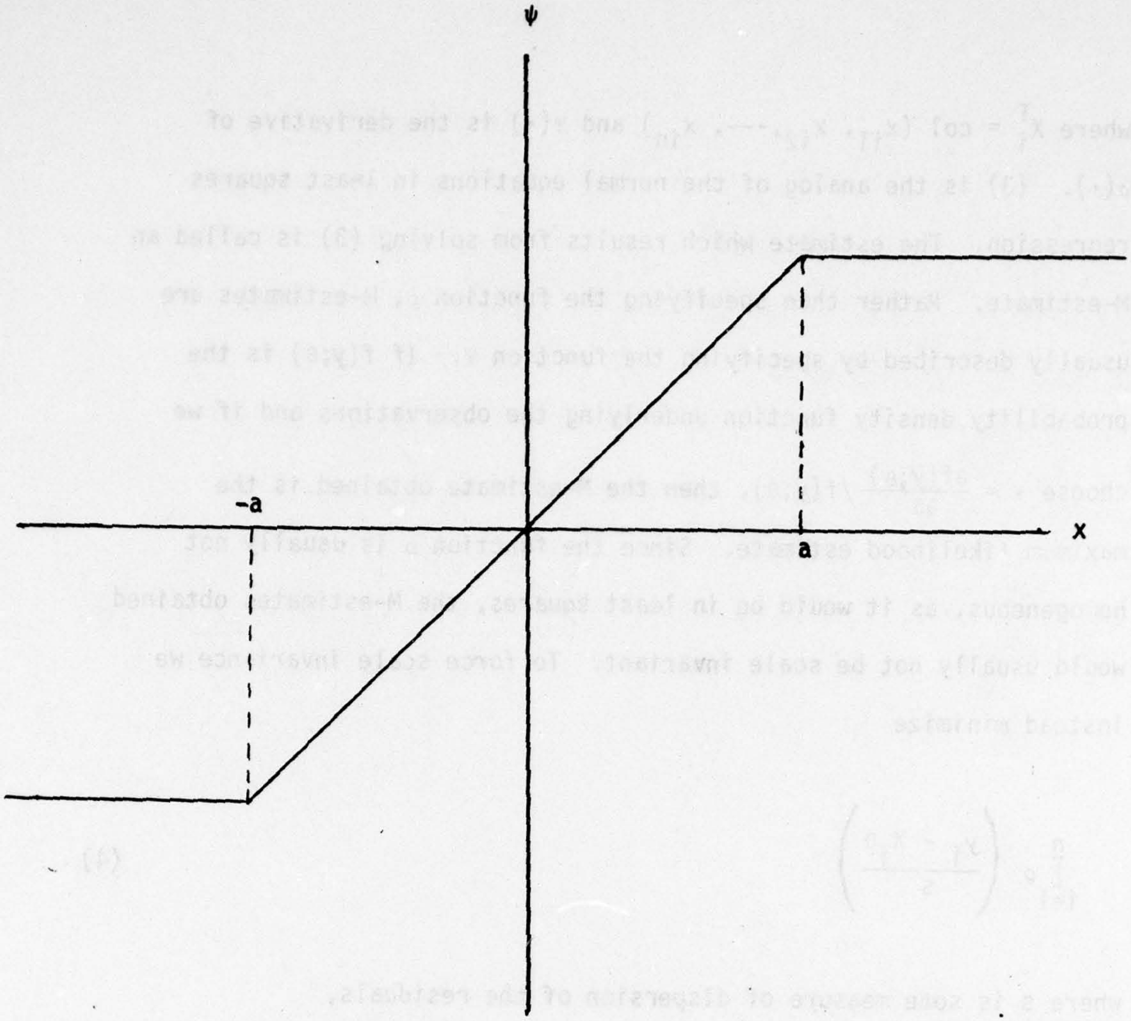
where $X_i^T = \text{col}\ (x_{i1}, x_{i2}, ---, x_{in})$ and $\Psi(\cdot)$ is the derivative of $\rho(\cdot)$. (3) is the analog of the normal equations in least squares regression. The estimate which results from solving (3) is called an M-estimate. Rather than specifying the function ρ, M-estimates are usually described by specifying the function Ψ. If $f(y;\theta)$ is the probability density function underlying the observations and if we choose $\Psi = \frac{\partial f(y;\theta)}{\partial\theta} / f(y;\theta)$, then the M-estimate obtained is the maximum likelihood estimate. Since the function ρ is usually not homogeneous, as it would be in least squares, the M-estimates obtained would usually not be scale invariant. To force scale invariance we instead minimize

$$\sum_{i=1}^{n} \rho \left(\frac{y_i - X_i\theta}{s} \right) \tag{4}$$

where s is some measure of dispersion of the residuals, $y_i - \sum_{j=1}^{p} x_{ij}\theta_j$. The quantity s also needs to be robust.

Several Ψ functions have been proposed in the literature. The original Ψ function proposed by Huber limits the sensitivity of the estimator to gross errors in the data. This Ψ function is given by

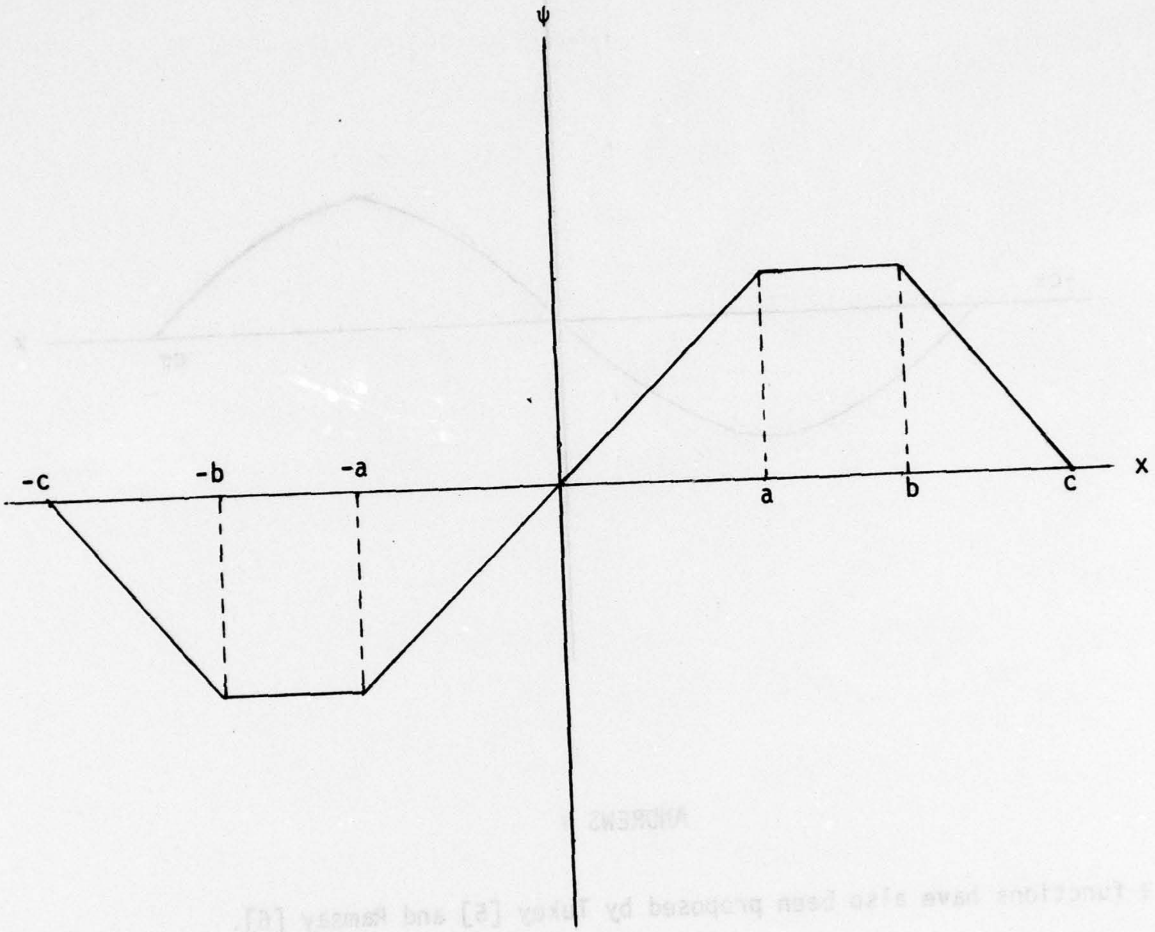
$$\Psi(x) = \begin{cases} x & |x| \leq a \\ a\ \text{sgn}\ (x) & |x| > a \end{cases}$$



HUBER - Ψ

A Ψ function of a different type which is an example of the redescending type of Ψ function and which provides rejection of gross errors as well as limited error sensitivity is the function proposed by Hampel [3].

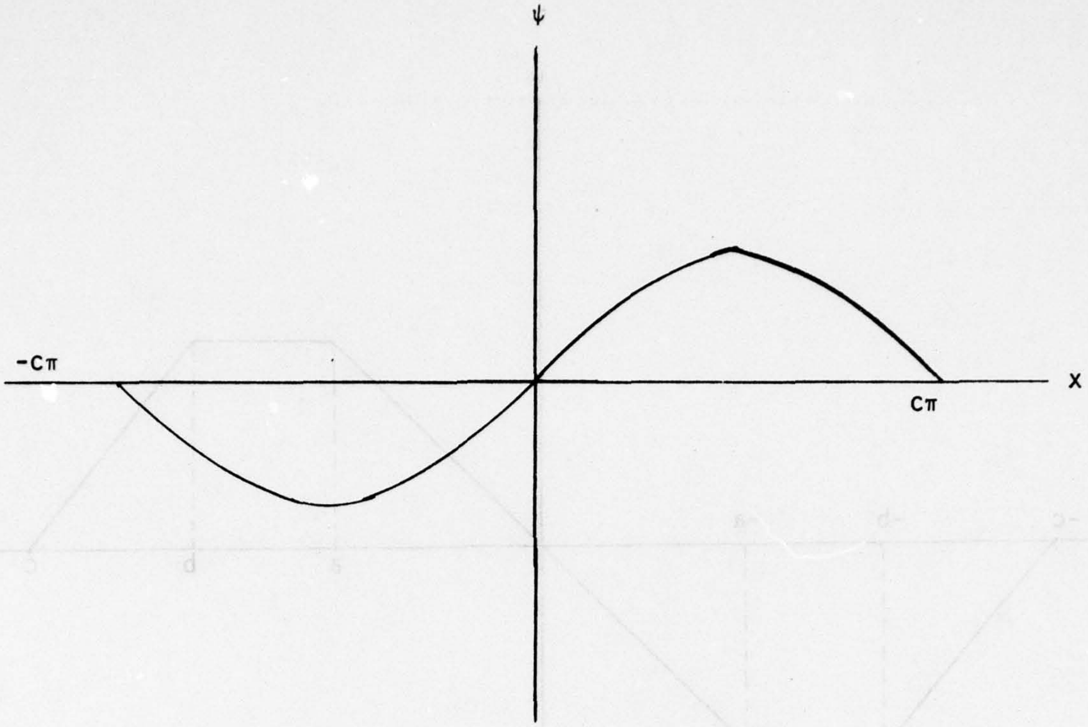
$$\Psi(x) = \begin{cases} x & |x| \leq a \\ a \operatorname{sgn}(x) & a < |x| \leq b \\ a\left(\dfrac{x - c \operatorname{sgn}(x)}{b - c}\right) & b < |x| \leq c \\ 0 & |x| \geq c \end{cases}$$



Hampel Ψ

A Ψ function proposed by Andrews [4] is given by

$$\Psi(x) = \begin{cases} \sin\left(\frac{x}{c}\right) & |x| \leq c\pi \\ 0 & |x| > c\pi \end{cases}$$



ANDREWS Ψ

Ψ functions have also been proposed by Tukey [5] and Ramsay [6].

Tukey's Ψ function is given by

$$\Psi(x) = \begin{cases} x\left(1 - \frac{x^2}{a^2}\right)^2 & |x|<a \\ 0 & |x|\geq a \end{cases}$$

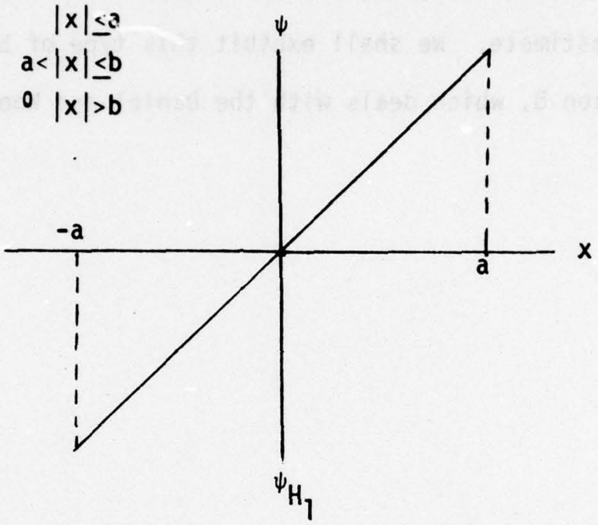
Another Ψ function which was proposed by Ramsay is

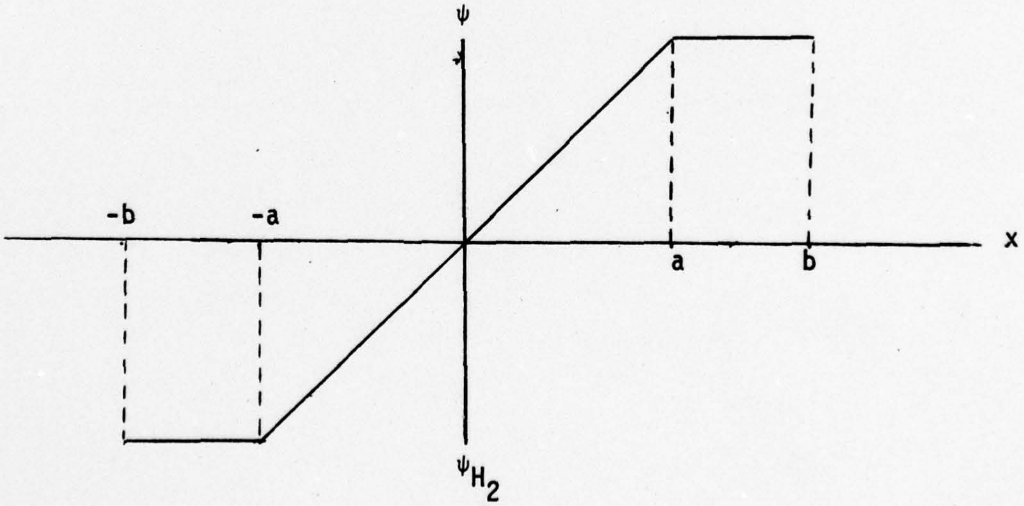
$$\Psi(x) = xe^{-a|x|}$$

The Ramsay ψ is of the redescending variety but the descent is very slow in comparison with other redescending ψ functions. Two ψ functions which are special cases of the Hampel ψ which we have found to be useful are for a=b=c and b=c. We will call these H_1 and H_2, respectively. They are given by

$$\psi_{H_1}(x) = \begin{cases} x & |x| \le a \\ 0 & |x| > a \end{cases}$$

$$\psi_{H_2}(x) = \begin{cases} x & |x| \le a \\ a & a < |x| \le b \\ 0 & |x| > b \end{cases}$$





Estimates with bounded Ψ function tend to be robust. If the Ψ function also returns to zero the estimator will tend to reject the more gross outliers and will be robust for a larger proportion of outliers. However, the ρ functions corresponding to the Ψ functions of the redescending class are not convex. Therefore, the numerical solution for M-estimates using a redescending Ψ function may result in an estimate which does not correspond to a global minimum of (4). This convergence to the wrong estimate may result in a degraded robust estimate. We shall exhibit this type of behavior for a Hampel Ψ in Section 8, which deals with the Daniel and Wood data.

As an example of the ability of M-estimates to detect outliers consider the data set below which is a time sequence of real angular measurement data and contains some gross outliers which are obvious by inspection. A quadratic curve is fit to the data for the purpose of determining the outliers.

Observations	Residuals from Least Squares Fit	Residuals from Robust Fit
-1.70987	-.157774	.000012
-1.70942	-.000204	.000004
-1.70893	.105480	.000003
-1.70845	.159227	-.000015
-1.70793	.161087	-.000010
-1.70741	.111021	-.000021
-1.70682	.009099	.000022
-1.70626	-.144780	.000019
-1.70571	-.350595	-.000010
-1.70510	-.608277	.000005
-1.70449	-.917885	.000004
1.43777	1.862231	3.141637
1.44602	1.456410	3.149243
-1.70257	-2.158177	-.000007
1.44667	.473139	3.146558

The residuals from the ordinary least squares fit do not yield any information about the outliers in the data whereas the outliers among the residuals from the robust M-estimate are obvious. The robust M-estimate for this example used a Hampel Ψ with breakpoints 3, 6, 9.

Another example which has residuals in all regions of the Hampel Ψ-function is the following data set.

	LEAST SQUARES RESIDUALS	ROBUST RESIDUALS	OBSERVATION	NORMALIZED ROBUST RESIDUALS
1	-.011022	-.000278	.20642275	1.005559
2	-.009071	-.000006	.20973521	.020803
3	-.007471	-.000033	.21296912	.120171
4	-.005711	.000151	.21663652	.546808
5	-.004461	-.000123	.22006619	.445501
6	-.002730	.000136	.22425138	.492246
7	-.001590	-.000144	.22811853	.519552
8	-.000213	.000135	.23249603	.487267
9	.001201	-.000038	.23718297	.136926
10	.002489	-.000014	.24201791	.051970
11	.003798	.000082	.24714760	.297949
12	.005624	.000748	.25306741	2.703007
13	.005421	-.000564	.25723122	2.037977
14	.008660	.001617	.26510980	5.845255
15	.006010	-.002037	.26737381	7.361710
16	.009663	.000662	.27621340	2.394020
17	.011016	.001113	.28302583	4.023731
18	.010359	-.000392	.28810282	1.418292
19	.011568	.000019	.29531815	.067036
20	.012005	-.000291	.30203451	1.051051
21	.012861	-.000129	.30944403	.464413
22	.013557	-.000075	.31696650	.269818
23	.014001	-.000222	.32450959	.800901
24	.014501	-.000260	.33238295	.938668
25	.015039	-.000209	.34056693	.754131
26	.015433	-.000249	.34888132	.898506
27	.015913	-.000152	.35755414	.547835
28	.016283	-.000113	.36639033	.406971
29	.016494	-.000181	.37534057	.654202
30	-.058265	-.075167	.30959446	271.639732
31	-.172487	-.189565	.20465789	685.052254
32	.018472	.001270	.40517605	4.589248
33	-.064416	-.081690	.33212063	295.211357
34	.089274	.071980	.49591643	260.122231
35	-.251831	-.269092	.16519139	972.446930
36	.007852	-.009326	.43552655	33.701152
37	.159606	.142564	.59820610	515.197899
38	.059168	.042313	.50896735	152.912771
39	.016704	.000088	.47797510	.318960
40	.016296	-.000028	.48931307	.101770

The solution for the M-estimate used a least square starting solution and a Hampel Ψ function with breakpoints at 2.5, 5, and 7.5. In the list of least squares residuals given above some of the outliers are obvious while others are not. The column of normalized residuals is merely the robust residual divided by the robust dispersion measure s. If we declare that residuals greater than 2.5 s are outliers then we would flag observations 12, 14, 15, 17, 30, 31, 32, 33, 34, 35, 36, 37, and 38 as outliers. Some of these outliers are much more gross than others. The M-estimate of the parameter vector is $\hat{\theta}_0 = .20388$, $\hat{\theta}_1 = .05419$, $\hat{\theta}_2 = .04427$. This example is simulated data so that the true parameter vector is known to be $\theta_0 = .20388$, $\theta_1 = .0537$, $\theta_2 = .0445$. The least squares starting solution was $\theta_0^{(0)} = .21636$, $\theta_1^{(0)} = .01901$, $\theta_2^{(0)} = .05466$.

3. Numerical Computation of M-Estimates

One of the most attractive features of least squares estimation is the ease of numerical solution. One might be inclined to think that the numerical solution for M-estimates would in many cases be prohibitive. This is not the case. At worst (4) can be minimized by one of the many algorithms for minimization, e.g., the Fletcher - Powell [7]. However, either a Gauss-Newton or a weighted least squares solution can usually be applied to obtain the M-estimate.

The Gauss-Newton method can be applied to the computation of M-estimates by linearization of (4) or (5) below. Setting the derivative of (4) with respect to θ equal to zero

$$\sum_{i=1}^{N} X_i^T \psi\left(\frac{y_i - X_i\hat{\theta}}{s}\right) = 0 \qquad (5)$$

Since (5) is in general nonlinear in $\hat{\theta}$, we must usually employ some form of iteration for solution. Suppose we have obtained an estimate $\hat{\theta}^{(k)}$ in the iteration sequence. We will discuss methods for obtaining a starting solution $\hat{\theta}^{(0)}$ in a later section. Linearizing (5) about $\theta^{(k)}$

$$\sum_{i=1}^{N} X_i^T \left(\psi\left(\frac{r_i^{(k)}}{s}\right) - \frac{1}{s}\,\psi'\left(\frac{r_i^{(k)}}{s}\right) X_i \left(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}\right)\right) = 0 \qquad (6)$$

where $r_i^{(k)} = y_i - X_i\hat{\theta}^{(k)}$

solving for $\hat{\theta}^{(k+1)}$

$$\hat{\theta}^{(k+1)} = \theta^{(k)} + M^{-1} \sum_{i=1}^{N} \psi\left(\frac{r_i^{(k)}}{s}\right) X_i^T \qquad (7)$$

where

$$M = \sum_{i=1}^{N} \psi'\left(\frac{r_i^{(k)}}{s}\right) \frac{X_i^T X_i}{s} \qquad (8)$$

(7) and (8) are iterated until $||\theta^{(k+1)} - \theta^{(k)}||$ is less than some prescribed value or for a fixed number of iterations.

A somewhat simpler method for solution is obtained by approximation of the Gauss-Newton method. Replacing $\Psi\left(\frac{r_i(k)}{s}\right)$ in the above equations by its sample mean

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + M^{-1} s^2 \frac{\sum_{i=1}^{N} \Psi(\frac{r_i(k)}{s})X_i^T}{\frac{1}{N}\sum_{i=1}^{N} \Psi(\frac{r_i(k)}{s})} \qquad (9)$$

where

$$M = \sum_{i=1}^{N} X_i^T X_i \qquad (10)$$

The advantage of this simplified method is that M and its inverse need to be calculated only once during the iteration procedure.

A simple method for the computation of M-estimates which has achieved considerable popularity is the iterative application of weighted least squares. We rewrite (5) as

$$\sum_{i=1}^{N} \frac{\Psi(\frac{y_i - X_i\hat{\theta}}{s})}{(\frac{y_i - X_i\hat{\theta}}{s})} X_i^T(y_i - X_i\hat{\theta}) = 0 \qquad (11)$$

Now let

$$W_i(\hat{\theta}) = \frac{\psi(\frac{y_i - X_i\hat{\theta}}{s})}{(\frac{y_i - X_i\hat{\theta}}{s})} \qquad (12)$$

Then (11) is

$$\sum_{i=1}^{N} W_i(\hat{\theta})X_i^T(y_i - X_i\hat{\theta}) = 0 \qquad (13)$$

(13) can be solved iteratively as follows. Let $\hat{\theta}^{(k)}$ be an arbitrary point in the iteration sequence. Then we approximate (13) by

$$\sum_{i=1}^{N} W_i(\hat{\theta}^{(k)})X_i^T(y_i - X_i\hat{\theta}^{(k+1)}) = 0 \qquad (14)$$

Solving (14) for $\hat{\theta}^{(k+1)}$

$$\hat{\theta}^{(k+1)} = (\sum_{j=1}^{N} W_j(\hat{\theta}^{(k)})X_j^TX_j)^{-1} \sum_{i=1}^{N} W_i(\hat{\theta}^{(k)})X_i^Ty_i \qquad (15)$$

Thus, we can use an ordinary weighted least squares algorithm iteratively to obtain the M-estimate.

Throughout the discussion of M-estimates we have used the dispersion measure s of the residuals without any consideration for its computation. Robust dispersion measures are often taken to be a multiple of the interquartile range or of some other range statistic. A dispersion measure which has been popular with those using M-estimates is the median deviation or the MAD (Median of the Absolute Deviations) estimate as it is sometimes called. The MAD estimate for regression is defined by

$$s = \operatorname*{med}_{i}|r_i| \Big/ .6745 \qquad (16)$$

where $r_i = y_i - X_i\theta$. Hampel [3] has shown that the MAD estimate is the most robust estimate of dispersion. In the iterative schemes described above a new value of s is computed at each stage of the iteration using the most recent set of residuals. Thus in obtaining an estimate $\hat{\theta}^{(k+1)}$ we use

$$s = \operatorname*{med}_{i}|r_i^{(k)}| \Big/ .6745 \qquad (17)$$

where $r_i^{(k)} = y_i - X_i\hat{\theta}^{(k)}$.

Testing of the Gauss-Newton and the weighted least squares methods for the computation M-estimates on the Daniel and Wood data, which is presented in a later section showed that the weighted least squares method to be far better than Gauss-Newton. The Gauss-Newton had very poor convergence properties for this data, especially when using the Andrews Ψ function.

4. Covariance of Estimates

An approximate covariance for an M-estimate can be obtained from the Gauss-Newton method. Assuming the observation errors e_i and e_j in (1) to be statistically independent we use (7) and (8) to obtain the approximate covariance for $\hat{\theta}$.

$$\mathrm{cov}(\hat{\theta}) \approx E\left[\Psi^2\left(\frac{y_i - X_i\theta}{s}\right)\right] M^{-1}\left(\sum_{j=1}^{N} X_j^T X_j\right) M^{-1} \qquad (18)$$

We further approximate $\mathrm{cov}(\hat{\theta})$ by replacing the expectation in (18) by its sample mean. Thus, we obtain

$$\mathrm{cov}(\hat{\theta}) \approx \frac{1}{n-p} \sum_{i=1}^{N} \Psi^2\left(\frac{y_i - X_i\hat{\theta}}{s}\right) M^{-1}\left(\sum_{j=1}^{N} X_j^T X_j\right) M^{-1} \qquad (19)$$

Corresponding to the approximation used to obtain (9) and (10) we can further approximate (19) by replacing $\psi'(\frac{r_i^0}{s})$ in M by its sample mean. Using this in (19) gives an alternative approximation to the covariance

$$\mathrm{cov}(\hat{\theta}) \approx \frac{\frac{1}{n-p} \sum_{j=1}^{N} \psi^2\left(\frac{y_j - X_j\hat{\theta}}{s}\right)}{\left[\frac{1}{N} \sum_{j=1}^{N} \psi'\left(\frac{y_j - X_j\hat{\theta}}{s}\right)\right]^2} \, s^2 \left(\sum_{i=1}^{N} X_i X_i^T \right)^{-1} \qquad (20)$$

In [1] Huber considers the asymptotic bias of the expressions (19) and (20). Huber also gives another alternative approximation to the covariance for an M-estimate.

5. Starting Solutions

Any of the numerical methods used to obtain an M-estimate requires a starting or preliminary estimate of the regression parameters θ. The starting solution is of primary importance and for some cases will determine whether or not a usable M-estimate is obtained. Robust estimation using Ψ functions of the redescending type is especially sensitive to the starting solution because the solution iteration may converge to a local minimum which is relatively remote from the global minimum, if a poor starting solution is used. At best, poor starting solutions require more iterations for convergence. The most obvious solution with which to start the M-estimation iteration is the unweighted least squares solution. However, since the unweighted least squares solution is highly influenced by the presence of outliers, it may not provide a suitable starting solution, $\hat{\theta}^{(0)}$. Nevertheless, least squares is often useful for starting. In some cases where the y_i are small and the components of θ are also small the starting solution $\hat{\theta}^{(0)} = 0$ may be useful. This is often the case in instrument calibration, see [8].

A good starting solution should itself be a robust estimate of the regression coefficients. Although the use of a robust starting solution may greatly increase the computing time, it will often be necessary if the two simple procedures mentioned above fail. Several robust regression methods which are suitable starting procedures for M-estimates are described in [9]. One of the simplest of these methods is an extension of the method proposed by Theil [10]. In applying this method we include a constant term θ_o separately from the other terms in the linear model. We then apply a Gram-Schmidt orthogonalization process to the remaining independent variables. The computation of the values X'_{ij} of the orthogonal variables is given by

$$X'_{i1} = X_{i1} \qquad (21)$$

$$X'_{ij} = X_{ij} - \sum_{k=1}^{j-1} r_{jk} X'_{ik} \qquad (22)$$

$$r_{jk} = \sum_{i=1}^{N} X_{ij} X'_{ik} \Big/ \sum_{i=1}^{N} X'^{2}_{ik} \qquad (23)$$

In terms of the orthogonal independent variables the linear model is given by

$$y_i = \theta_o + \sum_{j=1}^{p-1} X'_{ij} \theta'_j + e_i, \; i=1,N \qquad (24)$$

Estimates of the regression coefficients θ'_j are obtained using our modified method of Theil by the following process.

$$\left.\begin{array}{ll}
1. & d_m(i,j) = \dfrac{y_j - y_i}{X'_{jm} - X'_{im}} \quad j>i \quad i=1,N-1 \\[2ex]
2. & \delta\theta'_m = \underset{i,j}{\text{med}}\ d_m(i,j) \\[2ex]
3. & \theta'_m \leftarrow \theta'_m + \delta\theta'_m \\[2ex]
4, & y_i \leftarrow y_i - \delta\theta'_m X'_{im} \quad i=1,N
\end{array}\right\} m=1,p-1$$

5. Repeat steps 1-4 until convergence.

6. $\hat{\theta}_o = \underset{i}{\text{med}}\ y_i$

In the above $\underset{i}{\text{med}}\ z_i$ means to take the median of the variables z_i over the index set i. In order to recover the original regression coefficients, it is necessary to apply the Gram-Schmidt process to the θ'_j.

$$\theta_{p-1} = \theta'_{p-1} \qquad (25)$$

$$\theta_{p-1-i} = \theta'_{p-1-i} - \sum_{j=0}^{i-1} r_{p-1-j,p-1-i}\theta_{p-1-j} \quad i=1,p-2 \qquad (26)$$

For even moderate values of N the number of slopes $d_m(i,j)$ which must be computed is quite large. Rather than use all of these slopes we can instead work with a reduced number of slopes. One possible reduced set of slopes can be obtained letting the x'_{im} be arranged in increasing order for each m and let $N^* = [\frac{N+1}{2}]$. Thus, if N is odd x_{N^*m} is the median of the x'_{im}, i=1,N. We then use the slopes

$$d_m(i) = \frac{y_{N^*+i} - y_i}{x_{N^*+i,m} - x'_{im}} \qquad \begin{array}{ll} i=1,N^* & \text{(N even)} \\ i=1,N^* - 1 & \text{(N odd)} \end{array}$$

These slopes are then used in step 2 of the iteration process with

$$\delta\theta_m = \operatorname*{med}_{j} d_m(j).$$

Another robust regression method for obtaining a starting solution for M-estimates is an application of Spearmans ρ as described in [9]. We again form a set of orthogonal independent variables x'_{im} i=1,N by applying the Gram-Schmidt process in (21) - (23). Let $R_{x_{im}}$ be the rank of x'_{im} among the x'_{jm} j=1,N and let R_{y_i} be the rank of y_i among the y_j, j=1,N. Then Spearmans ρ, a nonparametric estimate of the population correlation coefficient is defined as

$$\rho_{x_m y} = \frac{\sum_{i=1}^{N} (R_{x_{im}} - \overline{R}_{x_m})(R_{y_i} - \overline{R}_y)}{\sqrt{\sum_{i=1}^{N} (R_{x_{im}} - \overline{R}_m)^2}} \qquad (27)$$

where $\overline{R}_{x_m} = \overline{R}_y = \frac{N+1}{2}$.

is just the ordinary defining equation for the correlation coefficient with the variates replaced by ranks. A more useful definition of $\rho_{x_m y}$ for computing is

$$\rho_{x_m y} = 1 - \frac{6 \sum_{i=1}^{N} d_i^{\,2}}{N(N^2 - 1)} \qquad (28)$$

where d_i is the rank difference

$$d_i = R_{y_i} - R_{x_{im}}$$

In an orthogonal regression model the estimates of the regression coefficients may be written as

$$\hat{\theta}'_m = \hat{\rho}_{x_m y} \frac{\hat{\sigma}_y}{\hat{\sigma}_{x_m}} \tag{29}$$

where $\hat{\rho}_{x_m y}$, $\hat{\sigma}_y$, $\hat{\sigma}_{x_m}$ are the usual sample correlation coefficient and standard deviations. An obvious method of obtaining a robust estimate of θ'_m is to replace $\hat{\rho}_{x_m y}$, $\hat{\sigma}_y$, $\hat{\sigma}_{x_m}$ in (29) by nonparametric estimates of these quantities. Thus, we replace $\hat{\rho}_{x_m y}$ by Spearmans ρ and replace $\hat{\sigma}_y$ by

$$\hat{\sigma}_y = \frac{\underset{i}{\text{med}} |y_i - y^*|}{.6745} \tag{30}$$

where $y^* = \underset{i}{\text{med}}\, y_i$. We could also replace $\hat{\sigma}_{x_m}$ by an estimate similar to (30) but in most cases $\hat{\sigma}_{x_m}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x'_{im} - \overline{x}_m)^2$ is sufficient. The process is used iteratively to improve the estimate of θ'_m. The procedure is implemented by the following steps.

1. $R_{x_{im}}$ = rank x'_{im}

$$\hat{\sigma}_{x_m} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x'_{im} - \bar{x}_m)^2}$$

$$\left.\begin{aligned}
&\text{2.}\quad R_{y_i} = \text{rank } y_i \\
&\qquad y^* = \underset{i}{\text{med}}\ y_i \\
&\qquad \hat{\sigma}_y = \frac{\underset{i}{\text{med}} |y_i - y^*|}{.6745} \\
&\text{3.}\quad d_i = R_{y_i} - R_{x_{im}} \\
&\qquad \delta\rho_m = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)} \\
&\qquad \delta\theta'_m = \delta\rho_m \frac{\hat{\sigma}_y}{\hat{\sigma}_{x_m}} \\
&\qquad \theta'_m \leftarrow \theta'_m + \delta\theta'_m \\
&\qquad y_i \leftarrow y_i - \delta\theta'_m x'_{im} \quad i=1,N
\end{aligned}\right\} m=1,p-1$$

4. Repeat steps 2-3 until convergence.

5. $\hat{\theta}_o = \underset{i}{\text{med}}\ y_i$

As before we must apply the Gram-Schmidt process to the θ'_m in order to recover the original regression coefficients.

A third method for obtaining a robust starting solution is the orthogonal Brown-Mood method. This is a variation of the Brown-Mood method [11] which uses orthogonal independent variables. Let x'_{im}, m=1,p-1,i=1,N be a set of orthogonal independent variables obtained by applying the Gram-Schmidt process. Let x^*_m be the median of the x'_{im} i=1,N. The Brown-Mood method is iterative so let $\hat{\theta}^{(k)}_m$ be some estimate in the iteration sequence and let $r^{(k)}_i$ be the residuals

$$r^{(k)}_i = y_i - \sum_{m=1}^{p-1} x'_{im}\ \hat{\theta}'^{(k)}_m \qquad (31)$$

The Brown-Mood method computes corrections $\delta\theta'_m$ to $\hat{\theta}_m^{(k)}$ by

$$\delta\theta'_m = \frac{r_i^{(k)^+} - r_i^{(k)^-}}{x_m^+ - x_m^-} \tag{32}$$

where

$$x_m^+ = \underset{i\varepsilon I_U}{\text{med}}\ x'_{im} \qquad I_U = \{i \mid x'_{im} > x_m^*\} \tag{33}$$

$$x_m^- = \underset{i\varepsilon I_L}{\text{med}}\ x'_{im} \qquad I_L = \{i \mid x'_{im} \leq x_m^*\} \tag{34}$$

$$r_i^{(k)^+} = \underset{i\varepsilon I_U}{\text{med}}\ r_i^{(k)} \tag{35}$$

$$r_i^{(k)^-} = \underset{i\varepsilon I_L}{\text{med}}\ r_i^{(k)} \tag{36}$$

The estimates are updated by $\hat{\theta}'^{(k+1)}_m \leftarrow \hat{\theta}'^{(k)}_m + \delta\theta'_m$ and the above procedure is iterated to convergence. Finally, the estimate of θ_o is obtained from

$$\hat{\theta}_o = \underset{i}{\text{med}}\ r_i^{(k)}$$

The orthogonal Brown-Mood method is implemented by the following steps (starting with $\hat{\theta}^{(0)} = 0$)

$$\left.\begin{array}{ll}
1. & x_m^{*} = \underset{i}{\text{med}}\; x'_{im} \\
2. & x_m^{+} = \underset{i\varepsilon I_U}{\text{med}}\; x'_{im} \\
 & x_m^{-} = \underset{i\varepsilon I_L}{\text{med}}\; x'_{im} \\
3. & y_i^{+} = \underset{i\varepsilon I_U}{\text{med}}\; y_i \\
 & y_i^{-} = \underset{i\varepsilon I_L}{\text{med}}\; y_i \\
4. & \delta\theta'_m = \dfrac{y_i^{+} - y_i^{-}}{x_m^{+} - x_m^{-}} \\
 & \theta'_m \leftarrow \theta'_m + \delta\theta'_m \\
5. & y_i \leftarrow y_i - \delta\theta'_m \, x'_{im} \quad i=1,N
\end{array}\right\} m=1,p-1$$

6. Repeat steps 3-5 until convergence

7. $\theta_o = \underset{i}{\text{med}}\; y_i$

6. Robust Regression with Vector Observations

The problem of linear regression with observations of more than one dependent variables is quite common. In this case we are given N observations of each dependent variable y_α, α=1,m. We denote these observations by $y_\alpha(i)$, i=1,N_α, α=1,m. The vector of parameters to be estimated is still denoted by θ. The observations are related to the parameter vector by the linear model

$$y(i) = A(i)\theta + e(i) \quad i=1,N \tag{37}$$

where A(i) is an mxp matrix, y(i) is an m-vector of observations and e(i) is an m-vector of measurement errors. A least squares estimate of θ would minimize

$$\sum_{i=1}^{N} (y(i) - A(i)\theta)^T \; (y(i) - A(i)\theta) \tag{38}$$

A robust alternative to the least squares estimate would minimize

$$\sum_{i=1}^{N} \sum_{\alpha=1}^{m} \rho_\alpha \left(\frac{y_\alpha(i) - a_\alpha(i)\theta}{s_\alpha} \right) \tag{39}$$

where a (i) is the αth row of A(i) and $\rho_\alpha(\cdot)$ may be a different function for each of the dependent variables, and s_α is a robust estimate of dispersion for the residual $y_\alpha(i) - a_\alpha(i)\theta$. Setting the derivative of (39) with respect to θ to zero gives

$$\sum_{i=1}^{N} \sum_{\alpha=1}^{m} \frac{a_\alpha^T(i)}{s_\alpha} \psi_\alpha \left(\frac{y_\alpha(i) - a_\alpha(i)\theta}{s_\alpha} \right) = 0 \tag{40}$$

(40) can be conveniently be rewritten as

$$\sum_{i=1}^{N} A^T(i)D^{-1}\,\Psi\left(D^{-1}(y(i) - A(i)\theta)\right) \qquad (41)$$

where D is the diagonal matrix D=diag $(s_1, s_2, ---, s_m)$ and Ψ is the vector of ψ functions

$$\Psi(x) = \begin{bmatrix} \psi_1(x_1) \\ \psi_2(x_2) \\ \cdot \\ \cdot \\ \cdot \\ \psi_m(x_m) \end{bmatrix} \qquad (42)$$

Either a Gauss-Newton or a weighted least squares solution can be used to iteratively obtain the M-estimate from (42). If $\hat{\theta}^{(k)}$ is an arbitrary point in the iteration sequence, the weighted least squares method applied to (42) gives

$$\hat{\theta}^{(k+1)} = M^{-1} \sum_{i=1}^{N} A^T(i)D^{-1}W(i)D^{-1}y(i) \qquad (43)$$

where D is the diagonal matrix

$$D = \text{diag } (s_1, s_2, ---, s_m)$$

and

$$M = \sum_{j=1}^{N} A^T(i)D^{-1}W(i)D^{-1}A(i) \qquad (44)$$

W(i) is a matrix of weights given by

$$W(i)=\text{diag}\left(\frac{\psi_1\left(\frac{r_1^{(k)}(i)}{s_1}\right)}{\frac{r_1^{(k)}(i)}{s_1}}, \frac{\psi_2\left(\frac{r_2^{(k)}(i)}{s_2}\right)}{\frac{r_2^{(k)}(i)}{s_2}}, --- \frac{\psi_m\left(\frac{r_m^{(k)}(i)}{s_m}\right)}{\frac{r_m^{(k)}(i)}{s_m}} \right) \qquad (45)$$

where $r_\alpha^{(k)}(i)$ is the residual

$$r_\alpha^{(k)}(i) = y_\alpha(i) - a_\alpha(i)\theta^{(k)} \tag{46}$$

As an example of robust linear regression with vector observations consider the calibration of a laser tracker. The laser tracker measures the range, azimuth and elevation of M targets with known range, azimuth, and elevation. Calibration constants for the tracker are computed by comparing the observations against the known positions of the M targets. Let R_{sj}, E_{sj}, and A_{sj} be the known surveyed range, azimuth, and elevation of the jth target. Suppose that multiple observations of the targets are available so that we have N_j observations for the jth target. Denote these range, azimuth, and elevation observations by R_{ij}, A_{ij}, and E_{ij}, i=1, N_j, j=1,M. Let

$$\Delta R_{ij} = R_{ij} - R_{sj} = r_j^T\theta + r_{ij}$$

$$\Delta A_{ij} = A_{ij} - A_{sj} = a_j^T\theta + a_{ij}$$

$$\Delta E_{ij} = E_{ij} - E_{sj} = e_j^T\theta + e_{ij}$$

where θ is an unknown parameter vector, r_j, a_j, and e_j are known vectors, and r_{ij}, a_{ij}, e_{ij} are random error terms. A common model for r_j, a_j, and e_j is given by

$$r_j^T\theta = \theta_1 + \theta_2 R_{sj} \tag{47}$$

$$a_j^T\theta = \theta_3 - \theta_4 \tan E_{sj} \cos A_{sj} - \theta_5 \tan E_{sj} \sin A_{sj} - \theta_6 \cos E_{sj} \tag{48}$$

$$e_j^T\theta = \theta_7 + \theta_4 \sin A_{sj} - \theta_5 \cos A_{sj} \tag{49}$$

The M-estimate for this example minimizes

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\left[\rho\left(\frac{\Delta R_{ij}-r_j^T\theta}{s_r}\right)+\rho\left(\frac{\Delta A_{ij}-a_j^T\theta}{s_a}\right)+\rho\left(\frac{\Delta E_{ij}-e_j^T\theta}{s_e}\right)\right] \qquad (50)$$

where s_r, s_a, s_e are robust measures of the dispersion of the range, azimuth, and elevation residuals. Differentiating (50) gives the analog to the normal equations

$$\sum_{j=1}^{M}\sum_{i=1}^{N_j}\left[\psi\left(\frac{\Delta R_{ij}-r_j^T\hat{\theta}}{s_r}\right)\frac{r_j}{s_r}+\psi\left(\frac{\Delta A_{ij}-a_j^T\hat{\theta}}{s_a}\right)\frac{a_j}{s_a}+\psi\left(\frac{\Delta E_{ij}-e_j^T\hat{\theta}}{s_e}\right)\frac{e_j}{s_e}\right]=0 \qquad (51)$$

(51) is solved iteratively using the weighted least squares algorithm with

$$s_r = \operatorname*{med}_{i,j}|d_r(i,j)| \Big/ .6745$$

$$s_a = \operatorname*{med}_{i,j}|d_a(i,j)| \Big/ .6745$$

$$s_e = \operatorname*{med}_{i,j}|d_e(i,j)| \Big/ .6745$$

where

$$d_r(i,j) = \Delta R_{ij} - r_j^T\hat{\theta}$$

$$d_a(i,j) = \Delta A_{ij} - a_j^T\hat{\theta}$$

$$d_e(i,j) = \Delta E_{ij} - e_j^T\hat{\theta} \; .$$

The following illustrates the application of the above to real field data. The laser tracker is calibrated using range, azimuth, and elevation measurements from eight reflective, surveyed targets arranged in a circular pattern around the tracker at a range of about 2500 feet. We use the model in (47) - (49). Since the elevations of the eight targets are approximately equal, it is obviously impossible to estimate θ_6 in (48) without additional observations. In order to provide these extra observations, we observe the same calibration targets but with the tracker "dumped", i.e. with an azimuth of approximately A_{si} + 180° and an elevation of approximately E_{si} - 180°. These additional observations are called dumped readings and are treated as additional calibration targets. Also, it will not be possible to estimate θ_2 in (47) since all ranges are approximately equal. In order to estimate θ_2, we observe four additional targets with ranges varying from 20,000 feet to 60,000 feet. Robust estimation of θ was done for this example using a Hampel ψ function with breakpoints a=2.5, b=5.0, and c=7.5. Approximately 250 observations are available for each target. The results of this robust calibration are summarized in the following table by tabulating the number of residuals for each target lying in each region of the Hampel ψ. The number of residuals in each region is the sum of the number in the positive and corresponding negative regions of the ψ function. The first eight target boards are at 2500 ft. circularly about the tracker. Targets 9-12 are the long range target boards. Targets 13-20 are "dumped" readings of the first eight targets. From the table it is obvious that most of the observations from several target boards are outliers, particularly for the "dumped" readings. This example has about 22% contamination by outliers which is extreme for this application, but illustrates the power of the M-estimation process in dealing with many outliers.

NUMBER OF RESIDUALS

TARGET POLE #	$<2.5\ s_r$	$(2.5\ s_r, 5\ s_r)$	$(5\ s_r, 7.5\ s_r)$	$>7.5\ s_r$	$<2.5\ s_a$	$(2.5\ s_a, 5\ s_a)$	$(5\ s_a, 7.5\ s_a)$	$>7.5\ s_a$	$<2.5\ s_e$	$(2.5\ s_e, 5\ s_e)$	$(5\ s_e, 7.5\ s_e)$	$>7.5\ s_e$
1	230	0	0	3	230	0	0	3	230	0	0	3
2	252	0	0	0	251	0	0	1	252	0	0	0
3	237	0	0	0	237	0	0	0	237	0	0	0
4	270	0	0	0	270	0	0	0	270	0	0	0
5	241	1	0	0	242	0	0	0	242	0	0	0
6	242	0	0	0	242	0	0	0	242	0	0	0
7	237	0	0	0	237	0	0	0	237	0	0	0
8	215	0	0	9	215	0	0	9	222	2	0	0
9	9	0	0	241	7	2	0	241	193	57	0	0
10	269	15	0	0	243	40	1	0	284	0	0	0
11	250	1	0	0	245	6	0	0	251	0	0	0
12	161	2	0	118	127	35	55	64	217	0	0	64
13	118	103	0	5	224	0	0	2	38	186	0	2
14	135	86	25	4	248	0	0	2	234	13	0	3
15	2	0	0	0	2	0	0	0	2	0	0	0
16	126	59	9	33	221	0	0	6	21	200	0	6
17	138	96	20	39	248	0	0	45	248	44	1	0
18	2	0	0	0	0	0	0	2	0	2	0	0
19	137	71	9	17	0	0	0	234	8	226	0	0
20	81	86	18	111	0	0	0	296	296	0	0	0

RANGE AZIMUTH ELEVATION

DISTRIBUTION OF CALIBRATION RESIDUALS

7. Nonlinear Regression

Instead of estimating regression parameters in the linear model suppose we want an M-estimate of the parameter vector θ in the nonlinear model

$$y_i = f_i(\theta) + e_i, \ i=1,N \tag{52}$$

where $f_i(\cdot)$ is a given nonlinear function. Then an M-estimate of θ is obtained by minimizing

$$\sum_{i=1}^{N} \rho\left(\frac{y_i - f_i(\theta)}{s}\right) \tag{53}$$

Differentiating (53) with respect to θ gives the nonlinear equations

$$\sum_{i=1}^{N} F_i^T(\hat{\theta})\ \psi\left(\frac{y_i - f_i(\hat{\theta})}{s}\right) = 0 \tag{54}$$

where $F_i(\hat{\theta})$ is the derivative vector

$$F_i(\hat{\theta}) = \frac{\partial f_i(\theta)}{\partial\theta}\Bigg|_{\theta=\hat{\theta}} \tag{55}$$

(54) can be solved by iteration. Either Gauss-Newton or weighted least squares iteration can be used to solve (54). Suppose we use weighted least squares. We rewrite (54) as

$$\sum_{i=1}^{N} \frac{\psi\left(\frac{y_i - f_i(\hat{\theta})}{s}\right)}{\left(\frac{y_i - f_i(\theta)}{s}\right)}\ F_i^T(\hat{\theta}) \left(\frac{y_i - f_i(\hat{\theta})}{s}\right) = 0 \tag{56}$$

Let $\hat{\theta}^{(k)}$ be an arbitrary point in the iteration sequence. Linearizing (56) about $\hat{\theta}^{(k)}$ and discarding higher order terms gives

$$\sum_{i=1}^{N} W_i(\hat{\theta}^{(k)})\, F_i^T(\hat{\theta}^{(k)}) \left(y_i - F_i(\hat{\theta}^{(k)})(\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}) \right) = 0 \qquad (57)$$

where

$$W_i(\hat{\theta}^{(k)}) = \frac{\psi\left(\dfrac{y_i - f_i(\hat{\theta}^{(k)})}{s}\right)}{\dfrac{y_i - f_i(\hat{\theta}^{(k)})}{s}} \qquad (58)$$

Solving (57) for $\hat{\theta}^{(k+1)}$

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \left(\sum_{j=1}^{N} W_j(\hat{\theta}^{(k)}) F_j^T(\hat{\theta}^{(k)}) F_j(\hat{\theta}^{(k)}) \right)^{-1} \sum_{i=1}^{N} W_i(\hat{\theta}^{(k)}) F_i^T(\hat{\theta}^{(k)}) y_i \qquad (59)$$

The choice of starting solution for a nonlinear problem presents additional difficulty if the unweighted least squares solution is not suitable. Methods for obtaining other starting solutions are dependent on the nature of the problem.

As an example of the application of M-estimates with a nonlinear model consider the N-station cinetheodolite trajectory data reduction problem. In this situation we are given azimuth observations $a_\alpha(t_i)$ and elevation observations $e_\alpha(t_i)$, $\alpha=1,N_i$ at each time point t_i along a trajectory. From these N_i cinetheodolites (tracking cameras) we must estimate the cartesian position $x(t_i)$, $y(t_i)$, $z(t_i)$ at each time point. The observations are $a_\alpha(t_i)=A_\alpha(\overline{x}_i)+$ error and $e_\alpha(t_i)=E_\alpha(\overline{x}_i)+$ error, where $\overline{x}_i$ is the 3-vector $[x(t_i)\ y(t_i)\ z(t_i)]$. The measurement functions $A_\alpha(\overline{x}_i)$ and $E_\alpha(\overline{x}_i)$ are given by

$$A_\alpha(\overline{x}_i) = \tan^{-1}\frac{x(t_i) - x_\alpha}{y(t_i) - y_\alpha}$$

$$E_\alpha(\overline{x}_i) = \tan^{-1}\frac{z(t_i) - z_\alpha}{[(x(t_i) - x_\alpha)^2+(y(t_i) - y_\alpha)^2]^{1/2}}$$

where $(x_\alpha, y_\alpha, z_\alpha)$ is the cartesian position of the αth camera. In this case we have a nonlinear regression problem with vector observations. In this application we minimize

$$\sum_{\alpha=1}^{N_i}\left[\rho\left(\frac{a_\alpha(t_i)-A_\alpha(\overline{x}_i)}{s_a}\right)\cos^2 e_\alpha(t_i)+\rho\left(\frac{e_\alpha(t_i)-E_\alpha(\overline{x}_i)}{s_e}\right)\right]$$

As a numerical example of this application consider the following situation which is rather extreme but sometimes occurs. A missile is fired at a drone and cinetheodolites are observing both the missile and the drone. It is required to estimate trajectories for both the missile and drone. Due to an inadvertent clerical error, one of the cameras which was actually observing the missile was erroneously listed as observing the drone. Obviously, when doing a least squares solution to obtain the drone trajectory, the azimuths and elevations from one camera will be gross outliers and will destroy the least squares solution for the drone position coordinates. A single point of this situation is given by the data below.

Camera	Obs. Azimuth	Obs. Elevation
1	.568106	.338886
2	-.626010	.122620
3	-2.665036	.359168
4	1.926249	.327177

Camera 2 is the one which is actually tracking the missile rather than the drone. Obviously, as in most situations which are the nonlinear, there is no way of distinguishing the outliers by inspecting the observations. As always in robust estimation a preliminary solution is required to start the iteration. Let $(x_\alpha, y_\alpha, z_\alpha)$ be a position solution obtained from the αth pair of cameras. In this example we have six possible pairs of cameras so that $\alpha=1,6$. We then start the iteration with $(x^\circ, y^\circ, z^\circ)$ where $x^\circ=$ med x_α, $\alpha=1,6$, $y^\circ=$ med y_α, $\alpha=1,6$, $z^\circ=$ med z_α, $\alpha=1,6$. For the example, the median guess solution is $x^\circ= -45147.9$ ft., $y^\circ= 87423.8$ ft., $z^\circ= 11117.3$ ft. After five iterations the sequence has converged to the solution $x= 32964.8$ ft., $y= 87425.2$ ft., $z= 11114.9$ ft. The residuals from the final solution are

	RESIDUALS	
CAMERA	AZIMUTH	ELEVATION
1	.000008	-.000064
2	-.242553	.011513
3	.000022	.000081
4	.000057	-.000019

Thus, the robust solution using the Hampel ψ with breakpoints of 3, 6, 9, correctly identified the outliers. Let us carry this example farther. Suppose we have no observations from camera 1, i.e., we have data from only three cameras one of which is bad. In this case our starting solution turns out to be $x^\circ= 45147.9$, $y^\circ= 87424.1$, $z^\circ= 11120.2$. After four iterations the solution has converged to $x= -32966$, $y= 87424.6$, $z= 11115.3$. Thus, we are again able to correctly identify the bad camera. Now suppose we have data from cameras 1, 2, 3. In this, the initial guess solution is $x^\circ = 45147.9$, $y^\circ = 67033.9$, $z^\circ = 11118.9$. After ten iterations the solution is $x = -35023.9$, $y = 84462.1$, $z = 11004.1$. The solution eventually converges to the correct value, but slowly. A third possibility to have data from only three cameras is observations from cameras 1, 2, 4. In this case the guess solution is $x^\circ = -46454.3$, $y^\circ = 87548.3$, $z^\circ = 7262.7$. After three iterations the solution has converged to $x = -35392.6$, $y = 86464.3$, $z = 1044.8$. Thus, in this case the iteration has converged to the wrong solution. In the last two cases where the solution converged very slowly and converged to the wrong solution, the starting solution was too far from the correct solution. If a sufficiently good start had been provided, the solution would have converged correctly in a few iterations. If the number of cameras were great enough in comparison to the number of bad cameras, using the median of the solutions obtained from the camera pairs provides an acceptable starting solution. Unfortunately, the number of cameras is often no more than three or four. In the case of three cameras the use of a starting solution predicted from preceding points might be a desirable procedure.

8. EXAMPLE - The Daniel & Wood Data

The Daniel and Wood data has been used by several authors [4], [12], [13] to illustrate robust regression methods. The data is taken from Daniel and Wood [4], Chapter 5, where it is examined in considerable detail.

The Daniel and Wood data is a sequence of 21 observations in 3 independent variables given below

Obs #	y	x_1	x_2	x_3
1	42	80	27	89
2	37	80	27	88
3	37	75	25	90
4	28	62	24	87
5	18	62	22	87
6	18	62	23	87
7	19	62	24	93
8	20	62	24	93
9	15	58	23	87
10	14	58	18	80
11	14	58	18	89
12	13	58	17	88
13	11	58	18	82
14	12	58	19	93
15	8	50	18	89
16	7	50	18	86
17	8	50	19	72
18	8	50	19	79
19	9	50	20	80
20	15	56	20	82
21	15	70	20	91

The linear model assumed for this example is

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i} + e_i \quad i=1,21$$

The Daniel and Wood data is treated here first by ordinary least squares and then by M-estimates using Huber, Hampel, and Andrews ψ-functions combined with different possible starting solutions for these M-estimates. We denote the M-estimation process with a Huber ψ function having a breakpoint at x=a by $H_u(a)$, and with a Hampel ψ function having breakpoints of a, b, c by $H_a(a, b, c)$, and with an Andrews ψ function with parameter a by $A_n(a)$. When starting these M-estimation processes with the ordinary least squares solution, we obtain the following sets of regression parameter estimates.

	θ_0	θ_1	θ_2	θ_3
OLS	-39.92	.7156	1.295	-.1521
$H_u(1.4)$	-41.06	.8249	.9466	-.1291
$H_a(1.4,2.8,4.2)$	-42.88	.9233	.6736	-.1079
$A_n(1.4)$	-42.41	.9257	.6617	-.1120

The residuals from these solutions are

OBS \#	OLS	$H_u(1.4)$	$H_a(1.4,2.8,4.2)$	$A_n(1.4)$
1	3.23	3.01	2.43	2.46
2	-1.91	-2.12	-2.67	-2.65
3	4.56	4.16	3.50	3.52
4	5.70	6.44	6.86	6.88
5	-1.71	-1.67	-1.80	-1.79

OBS \#	OLS	$H_u(1.4)$	$H_a(1.4,2.8,4.2)$	$A_n(1.4)$
6	-3.01	-2.61	-2.47	-2.45
7	-2.39	-1.79	-1.50	-1.44
8	-1.39	- .79	- .50	- .44
9	-3.14	-2.31	-1.78	-1.75
10	1.27	.51	- .16	- .23
11	2.64	1.68	.81	.78
12	2.78	1.49	.37	.33
13	-1.42	-2.23	-2.95	-3.00
14	- .05	- .75	-1.43	-1.43
15	2.36	2.28	2.19	2.19
16	.90	.89	.87	.85
17	-1.59	- .87	- .31	- .38
18	- .46	.04	.44	.40
19	- .60	.22	.88	.85
20	1.41	1.53	1.55	1.52
21	-7.24	-8.86	-10.40	-10.43

In the above sets of residuals there are no grossly outlying observations so that we cannot readily judge the four regression methods. The robust methods have somewhat smaller residuals than the OLS method and possibly the regression with the Hampel or Andrews ψ-function gives slightly smaller residuals than regression with the Huber ψ-function. The nonparametric measure of dispersion for the residuals in each of the regressions is

	OLS	$H_u(1.4)$	$H_a(1.4,2.8,4.2)$	$A_n(1.4)$
s	2.83	2.49	2.30	2.25

If the residuals were tested for outliers against 3s, the OLS regression does not indicate any outliers, but the Huber, Hampel, and Andrews regressions indicate that the 21st observation is an outlier. In addition, the Andrews regression shows the 4th observation to be an outlier. Both the Hampel and Andrews regressions show the 21st observation as a gross outlier by giving it a zero weight.

Daniel and Wood, after some exhaustive analysis, declare that observations 1, 3, 4, and 21 are outliers. Also, in reading about the experiment from which the data were gathered, it is discovered that observations 1, 3, 4, and 21 were taken during transient conditions of the plant whereas the other observations were taken during steady state conditions. Thus, on the basis of Daniel and Woods work and the observations by the original experimenters observations 1, 3, 4, and 21 are probably outliers. The regression solution without these four points is $\hat{\theta}_0 = -37.65$, $\hat{\theta}_1 = .7977$, $\hat{\theta}_2 = .5773$, $\hat{\theta}_3 = -.0671$. The failure of the robust regressions to detect all of the outliers can be traced, at least in the case of the Hampel and Andrews regressions, to the inadequate least squares starting solution. We will demonstrate in the following that with a sufficiently good starting solution the Hampel and Andrews regressions will converge to solutions for which the outliers are obvious. Suppose we try the orthogonal Theil method, the Spearmans ρ method and the orthogonal Brown-Mood methods previously described to start the M-estimate regressions.

From these starting methods we obtain the following regression coefficients which will be used to start the M-estimates.

	θ_0	θ_1	θ_2	θ_3
Spearman ρ	-43.25	.7578	.8100	-.0257
Theil	-40.93	.7761	.6928	-.0384
Brown-Mood	-39.21	.7981	.3846	-.0000
OLS (w/o 1,3,4,21)	-37.65	.7977	.5773	-.0671

Both the H_a(1.4,2.8,4.2) and the A_n(1.4) regressions converge to the same solution as before when using the Spearman ρ starting solution. Also, the A_n(1.5) converges to the same solution as before when using the Theil starting solutions. The A_n(1.4) converges to a solution for which the outliers are obvious when using the OLS (w/o 1,3,4,21) or Brown-Mood starting solutions. Also, the H_a(1.4,2.8,4.2) regression converges to a solution for which the outliers are obvious when using either the Brown-Mood, the Theil or the OLS (w/o 1,3,4,21) starts. The regression coefficients obtained are

	θ_0	θ_1	θ_2	θ_3
A_n(1.4) from OLS (w/o 1,3,4,21) and Brown-Mood	-37.85	.8239	.5494	-.0751
H_a(1.4,2.8,4.2) from Brown-Mood, OLS (w/o 1,3,4,21) and Theil	-37.39	.8113	.5548	-.0734

The residuals from these solutions are

OBS #	A_n(1.4) from OLS (w/o 1,3,4,21) and Brown-Mood	H_a(1.4,2.8,4.2) from Brown-Mood and OLS (w/o 1,3,4,21)
1	5.78	6.04
2	.71	.97
3	6.08	6.28
4	8.11	8.16
5	- .79	- .73
6	-1.34	-1.28
7	- .44	- .40
8	.56	.60
9	-1.04	-1.04
10	.18	.22
11	.85	.88
12	.33	.37
13	-2.67	-2.63
14	-1.40	-1.38
15	1.44	1.38
16	.22	.15
17	- .38	- .43
18	.14	.09
19	.67	.60
20	1.88	1.88
21	-8.98	-8.81

The four outliers have now become fairly obvious among the residuals. Both of the regressions give zero weight to these observations. The dispersion measure for the residuals in Andrews regression is 1.26 and in the Hampel regression is 1.44.

The convergence of the $A_n(1.4)$ and $H_a(1.4,2.8,3.2)$ regressions on the Daniel and Wood data to a solution close to the OLS (w/o 1,3,4,21) regression in which the outliers are obvious is very sensitive to the starting solution. The sensitivity of the robust regressions to the starting solution for the Daniel and Wood data can be greatly lessened by changing the breakpoints of the ψ-functions so that we are doing $A_n(1)$ and $H_a(1,2,3)$ regressions. Both the $A_n(1)$ and $H_a(1,2,3)$ regressions converge to the same solutions starting from OLS, Spearman ρ, Theil, and Brown-Mood starting solutions. The regression coefficients obtained are

	θ_0	θ_1	θ_2	θ_3
$A_n(1)$	-37.11	.8190	.5175	-.0727
$H_a(1, 2, 3)$	-37.01	.8183	.5202	-.0742

The residuals from these solutions are

OBS #	$A_n(1)$	$H_a(1, 2, 3)$
1	6.09	6.11
2	1.02	1.04
3	6.30	6.32
4	8.24	8.25
5	- .72	- .71
6	-1.24	-1.23
7	- .32	- .30
8	.68	.70
9	- .96	- .96
10	.12	.13

OBS #	$A_n(1)$	$H_a(1, 2, 3)$
11	.77	.79
12	.21	.24
13	-2.74	-2.73
14	-1.46	-1.43
15	1.32	1.34
16	.10	.12
17	- .43	-.44
18	.08	.08
19	.63	.63
20	1.86	1.87
21	-8.95	-8.92

REFERENCES

1. Huber, Peter J., "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," Annals of Statistics, 1, (1973), p 799-821.

2. Huber, Peter J., "Robust Statistics: A Review," Annals of Mathematical Statistics, 43, (1972), p 1041-1067.

3. Hampel, Frank R., "The Influence Curve and its Role in Robust Estimation," Journal of the American Statistical Association, 69, (June 1974), p 383-393.

4. Andrews, D. F., "A Robust Method for Multiple Linear Regression," Technometrics, 16, (November 1974), p 523-531.

5. Tukey, John W., and Beaton, Albert E., "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," Technometrics, 16, (May 1974), p 147-192.

6. Ramsay, J. O., "A Comparitive Study of Several Robust Estimates of Slope, Intercept, and Scale in Linear Regression," Journal of the American Statistical Association, 72, (September 1977), p 608-615.

7. Fletcher, K., and Powell, M. J. D., "A Rapidly Convergent Descent Method for Minimization," Computer Journal, 6, 1963, p 163-168.

8. Agee, W. S., and Turner, R. H., "Application of Robust Statistical Methods to Data Reduction," Analysis and Computation Div., Tech. Rpt. No. 65, White Sands Missile Range, March 1978.

9. Agee, William S., and Turner, Robert H., "Robust Regression: Some New Methods and Improvements of Old Methods," Technical Report, White Sands Missile Range, 1978.

10. Theil, H., "A Rank-Invariant Method of Linear and Polynomial Regression Analysis," Indag. Math., 12, (1950), p 85-91, 173-177, 467-482.

11. Mood, A. M., "Introduction to the Theory of Statistics," McGraw-Hill, New York, 1950.

12. Hettmansperger, T. H., and McKean, J. W., "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models," Technometrics, 19, (August 1977), p 275-284.

13. Hogg, R. V., "Robust Statistical Procedures," Proceedings of the Twenty-Second Conference on the Design of Experiments in Army Research Development and Testing, (1977), p 247.

14. Daniel, C., and Wood, F. S., "Fitting Equations to Data," Chap. 5, Wiley-Interscience, New York, 1971.